A Conditioning Function for the Convergence of Numerical ODE Solvers and Lyapunov's Theory of Stability

Divakar Viswanath [*]

22 December 1998

Abstract

For the ordinary differential equation (ODE) $\dot{x}(t) = f(t,x)$, $x(0) = x_0$, $t \geq 0$, $x \in R^d$, assume f to be at least continuous in t and locally Lipshitz in x, and if necessary, several times continuously differentiable in t and x. We associate a conditioning function $E(t)$ with each solution $x(t)$ which captures the accumulation of global error in a numerical approximation in the following sense: if $\tilde{x}(t;h)$ is an approximation derived from a single step method of time step h and order r then $\|\tilde{x}(t;h) - x(t)\| < K(E(t) + \epsilon)h^r$ for $0 \leq t \leq T$, any $\epsilon > 0$, sufficiently small h, and a constant $K > 0$.

Using techniques from the stability theory of differential equations, this paper gives conditions on $x(t)$ for $E(t)$ to be upper bounded linearly or by a constant for $t \geq 0$. More concretely, these techniques give constant or linear bounds on $E(t)$ when $x(t)$ is a trajectory of a dynamical system which falls into a stable, hyperbolic fixed point; or into a stable, hyperbolic cycle; or into a normally hyperbolic and contracting manifold with quasiperiodic flow on the manifold.

1 Introduction

For the system of ordinary differential equations $\dot{x}(t) = f(t,x)$, $t \geq 0$, $x \in R^d$, the initial value problem $x(t_0) = x_0$, $t_0 \geq 0$, may not have a solution, and when the solution exists it may not be unique. However, if $f(t,x)$ is continuous in t and not only continuous but also locally Lipshitz in x, there is a unique solution of the initial value problem for $t_0 \leq t < t_0 + \epsilon$, $\epsilon > 0$, which we denote by $x(t;t_0,x_0)$. We make these assumptions about $f(t,x)$ and only consider solutions $x(t;t_0,x_0)$ that can be continued till $t = \infty$. Usually $t_0 = 0$, and we denote these solutions by $x(t;x_0)$ or simply $x(t)$.

Even at this level of generality, one might ask how accurately $x(t;x_0)$ can be approximated by a numerical method. A reasonable first guess is to look at the $d \times d$ matrix

$$E'(t) = \frac{\partial x(t;x_0)}{\partial x_0},$$

for $0 \leq t < \infty$; since $E'(t)$ gives the sensitivity of $x(t;x_0)$ with respect to x_0, one might hope to relate it to the accumulation of global error. However, numerical methods introduce discretization error not just at $t = 0$ but at every time step of the integration, and therefore, $E'(t)$ proves to be an insufficient concept.

[*] Research at MSRI is supported in part by NSF grant DMS-9701755

The following conditioning function $E(t)$ associated with $x(t;x_0)$, or more briefly $x(t)$, does capture the accumulation of global error:

$$E(t) = \sup_{v(s)} \Big\| \int_0^t \frac{\partial x(t)}{\partial x(s)} v(s) ds \Big\|, \tag{1.1}$$

with the supremum taken over continuous functions $v : [0,t] \to R^d$ with $\|v(s)\| \leq 1$ for $0 \leq s \leq t$. Unlike $E'(t)$, $E(t)$ takes into account the sensitivity of $x(t)$ with respect to all $x(s)$, $0 \leq s \leq t$. In the definitions of $E(t)$ and $E'(t)$ above, we have assumed $f(t,x)$ to be continuously differentiable with respect to t and x. All vector norms in this paper are Euclidean norms and all matrix norms are the corresponding induced norms.

Let $\tilde{x}(t;x_0;h)$ denote the approximation to $x(t;x_0)$ computed by a single step method of order r. The numerical method gives $\tilde{x}(t;x_0;h)$ at all positive integer multiples of h. For $kh < \tau < (k+1)h$, $k = 0,1,2,\ldots$, we define $\tilde{x}(\tau;x_0;h)$ by following the solution exactly from the initial point $t = kh$ and $x = \tilde{x}(kh;x_0;h)$ till $t = \tau$. Therefore, $\tilde{x}(t;x_0;h)$ can be discontinuous only at $t = kh$, $k = 1,2,\ldots$ The magnitude of this discontinuity

$$\|\tilde{x}((k+1)h;x_0;h) - x((k+1)h;kh,\tilde{x}(kh;x_0;h))\|$$

is the local discretization error at $t = (k+1)h$. We write this discretization error as $K_{k+1}h^{r+1}v_{k+1}$, where $v_{k+1} \in R^d$ with $\|v_{k+1}\| = 1$ and K_{k+1} is a non-negative real number. Thus the direction of the local discretization error at the ith step is v_i and its magnitude is $K_i h^{r+1}$.

Let us assume that the magnitude of the local discretization error is bounded above by Kh^{r+1} for a constant $K > 0$. This assumption can hold for example for a Runge-Kutta method of order r if $f(t,x)$ is $r+1$ times continuously differentiable with respect to t and x; we discuss this assumption further in Section 2. Then by Theorem 4.5, given $\epsilon > 0$ and $T > 0$,

$$\|x(t;x_0) - \tilde{x}(t;x_0;h)\| < (E(t) + \epsilon)Kh^r \tag{1.2}$$

for $0 \leq t \leq T$ and sufficiently small h. In this bound on the global error, $E(t)$, which is given by (1.1), is independent of details of the numerical method. Further, a bound like (1.2) will not hold for any real-valued function of t which is strictly less than $E(t)$ at some value of t. It is because of these reasons that we call $E(t)$ a *conditioning function.*

Let us now draw an analogy between $E(t)$ and the absolute condition number of a multivariate function $g(x)$. The formula (see [26])

$$\lim_{\delta \to 0^+} \sup_{\delta x < \delta} \frac{\|g(x+\delta x) - g(x)\|}{\delta}$$

is the analogue of (1.1); this absolute condition number measures the sensitivity of $g(x)$ with respect to small changes in x. For a stable numerical method, the error in evaluating $g(x)$ is governed by this absolute condition number in a manner similar to the dependence of global error on $E(t)$ given by (1.2).

Sections 2,3, and 4 define $E(t)$ and develop its properties. We do not begin with (1.1) as the definition of $E(t)$. Section 2 introduces a model of discretization errors which is similar to but more general than the model of Stuart and Humphries [25]. Section 3 defines $E(t)$ using this model. The expression for $E(t)$ in (1.1) is derived in Corollary 6.3 of Section 6. Let us mention the formal similarity of (1.1) to bounds on the global error derived by Iserles and Söderlind [16] using the Alexseev-Gröbner lemma. The results in this half of the paper argue that $E(t)$ is the appropriate vehicle for a study of global errors.

From Section 5 onwards, we undertake the task of relating $E(t)$ to stability properties of $x(t; x_0)$. The relation of $E(t)$ to the accumulation of global error as t increases is clear enough from (1.2). If $E(t)$ is bounded by a constant or linearly or by a polynomial of low degree in t, the accurate approximation of the trajectory $x(t; x_0)$ can be considered a tractable problem; but if $E(t)$ increases exponentially in t accurate approximation of $x(t; x_0)$ is pretty intractable. The standard technique of bounding global errors using the Lipshitz constant is used in convergence proofs of numerical methods [9]. But the bounds on $E(t)$ obtained this way increase exponentially in t and are of hardly any other use.

We upper bound $E(t)$ by making stability assumptions on $x(t; x_0)$. Sections 5 and 6 derive linear and constant bounds on $E(t)$ by making stability assumptions on $x(t; x_0)$. This connection to stability is somewhat subtle; there are exponentially stable examples with exponentially increasing $E(t)$. However, the work of researchers who followed Lyapunov allows us to clarify and circumvent the difficulties in relating $E(t)$ to stability theory. See Table 1 in Section 6 for a summary.

Let us now make some prefatory remarks about the stability theory of ordinary differential equations. The theory of stability of ordinary differential equations was initiated by A.M. Lyapunov in 1892 [17]. One stream of research which emanates from this remarkable work is about first approximations. Let $x(t) = x(t; x_0)$ be a solution of $\dot{x}(t) = f(t, x)$. If the perturbation $y(t)$ is such that $x(t) + y(t)$ is also a solution of the same equation, then $y(t)$ satisfies

$$\dot{y}(t) = f(t, y + x(t)) - f(t, x(t)) = F(t, y),$$

where clearly $F(t, 0) \equiv 0$. The solution $x(t)$ is stable in the sense of Lyapunov if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $\|y(0)\| < \delta$ implies $\|y(t)\| < \epsilon$ for $t \geq 0$. Now (see [22], [5]), $x(t)$ is stable if and only if the zero solution of $\dot{y}(t) = F(t, y)$ is stable. Thus it is enough to look at zero solutions of systems of the form $\dot{y}(t) = F(t, y)$, with $F(t, 0) \equiv 0$.

Lyapunov pointed out an other possible simplification. Since $F(t, 0) \equiv 0$, if $F(t, y)$ is assumed to be continuously differentiable in y, $F(t, y) = A(t)y + o(\|y\|)$ as $y \to 0$. Here,

$$\begin{aligned} A(t) &= \frac{\partial F(t, y)}{\partial y}\Big|_{y=0} \\ &= \frac{\partial f(t, x)}{\partial x}\Big|_{x=x(t;x_0)}. \end{aligned} \tag{1.3}$$

Lyapunov argued that perhaps stability properties of the zero solution of the linear first approximation $\dot{y}(t) = A(t)y$ might imply stability properties of the zero solution of the nonlinear system $\dot{y}(t) = F(t, y)$. This program has been carried out by Lyapunov, E. Cotton (1911), O. Perron (1928), I.G. Petrovskii (1934), R. Bellman (1953), and others.

The other stream of research, which also originated with Lyapunov, uses Lyapunov functions $V(t, y)$. If every solution $y(t; y_0)$ of $\dot{y}(t) = F(t, y)$, $t \geq 0$, had the property that $\|y(t; y_0)\|$ decreases as t increases, stability of the zero solution would be easy to infer. However, this property does not hold even for stable, linear systems of the form $\dot{y}(t) = Ay$, where A is a constant matrix. Lyapunov's idea was to use a non-negative, real-valued function $V(t, y)$, instead of the norm, such that $V(t, y(t))$ decreases as t increases. This $V(t, y)$ is required to be related to $\|y\|$ in a uniform way for $t \geq 0$; precise details depend upon the concept of stability. After Lyapunov, this line was greatly developed by Soviet researchers, including Persidskii, Malkin, Krasovskii, and others, beginning in the 1920s. It was taken up by researchers in the west, including J.L. Massera, T. Yoshizawa, and others, in the 1950s.

Parts of Sansone and Conti [22] and Hale [10] are excellent introductions to the theory of stability of ODEs. There are numerous advanced works; of them we refer to Bellman [1], Malkin [18], and Yoshizawa [27].

Obtaining detailed stability information about solutions of ODEs can be far from trivial. For this reason, applying the theory in Sections 5 and 6 to concrete examples is not a simple matter. In Section 7, we give three applications to dynamical systems. We derive constant or linear upper bounds on $E(t)$ for trajectories falling into stable, hyperbolic fixed points, or into stable, hyperbolic cycles, or into a normally hyperbolic and contracting manifold, with the flow on the manifold being quasiperiodic. These three applications involve the Hartman-Grobman theorem, convergence in phase results for stable cycles, and results of Pugh, Shub, and others about normally hyperbolic flows, respectively. The first of these applications was covered by Stuart and Humphries [25]. The second application is a significant extension of a result due to Cano and Sanz-Serna [4]. The third application appears to be entirely new.

Let us mention the similarity of our analysis to the asymptotic analysis of global error of Henrici [12] [13] and several following researchers, including Gragg [8]. Theorem 6.2, in particular, is implicit in Henrici's work. Beginning with Dahlquist [6], it has been known that the use of one-sided Lipshitz constants can sometimes give meaningful bounds on the global error. We show in Section 8 that the use of one-sided Lipshitz constants fits naturally into our framework. Section 8 also considers variable time stepping, multistep methods, and other issues.

To summarize briefly, this paper consists of three parts. The first part, Sections 2, 3, and 4, derives a conditioning function $E(t)$ and associates it with the global error in numerically approximating the solution of an ordinary differential equation. The second part, Sections 5 and 6, relates $E(t)$ to the stability theory of ordinary differential equations. The third part, Section 7, applies this theory to dynamical systems.

2 A Model for Discretization Errors

The model of discretization error which we now present is close to the way discretization errors are made by single step methods with constant step sizes. Stuart and Humphries [25] model discretization errors of single step methods in a similar manner.

Let $\alpha(h)$ be a continuous, strictly increasing function of h for $h \geq 0$. Assume also that $\alpha(0) = 0$. Then an approximation $\tilde{x}_\alpha(t; x_0; h)$ to $x(t; x_0)$ is defined as follows:

$$\begin{aligned} &\tilde{x}_\alpha(0; x_0; h) = x_0 \\ &\tilde{x}_\alpha(nh; x_0; h) = x(nh; (n-1)h, \tilde{x}_\alpha((n-1)h; x_0; h)) + h\alpha(h)v_n \quad n \geq 1, \end{aligned} \tag{2.1}$$

where $v_n \in R^d$ can be any vector with $\|v_n\| \leq 1$. In words, the approximate solution at $t = nh$, $n \geq 1$, is obtained by exactly propagating the point $\tilde{x}_\alpha((n-1)h; x_0; h)$ at $t = (n-1)h$ under $\dot{x}(t) = f(t, x)$ till $t = nh$, and then adding the discretization error or discontinuity $h\alpha(h)v_n$, where $\|v_n\| \leq 1$. For $(n-1)h \leq t < nh$,

$$\tilde{x}_\alpha(t; x_0; h) = x(t; (n-1)h, \tilde{x}_\alpha((n-1)h; x_0; h)). \tag{2.2}$$

Since v_n can be any vector with $\|v_n\| \leq 1$, this actually defines a whole family of approximate solutions which we denote by $\tilde{X}_\alpha(x_0; h)$. The exact solution $x(t)$ is the only member of this family which is continuous. Figure 1 gives an example of an approximate solution.

We now address how single step numerical methods are related to approximations from the family $\tilde{X}(x_0; h)$. In our description of single step methods, the discretization error at every step was taken to be $K_i h^{r+1} v_i$. We now assume that the K_i are bounded by a constant K which does not depend upon h or i. This can be proven in some circumstances; see [25]. Besides, if there is no such K, the numerical method will not in practice behave as if it were of order r. Thus

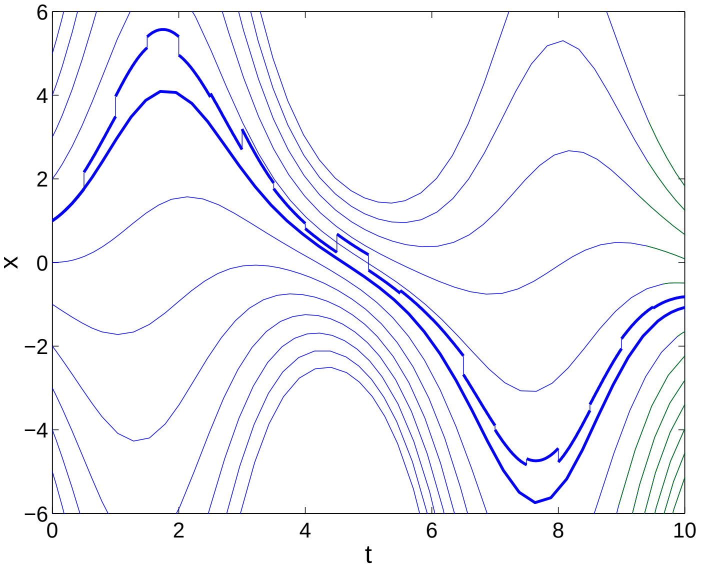

Figure 1: The thick lines show an exact solution of the equation $\dot{x}(t) = \sin t + (\cos t)x$ and an approximation to it with $h = 0.5$ and $\alpha(h) = 2h$ for.

when the order of accuracy of the numerical method for approximating $x(t; x_0)$ is r, we can take $\alpha(h) = Kh^r$, and there will always be an approximation in the family $\tilde{X}_\alpha(x_0; h)$ which is the same as the trajectory of the numerical method.

The notations we have established so far will be adhered to in the rest of the paper. For the system, $\dot{x}(t) = f(t, x)$, $f(t, 0)$ is not necessarily zero. The solution with $x(t_0) = x_0$ is denoted by $x(t; t_0, x_0)$, and by $x(t; x_0)$ when $t_0 = 0$. The approximations to $x(t; x_0)$ obtained as in (2.1) and (2.2) are denoted by $\tilde{x}_\alpha(t; x_0; h)$. The family of approximations is $\tilde{X}(x_0; h)$. The same notations apply for $\dot{y}(t) = F(t, y)$ with the xs changed to ys. But here $F(t, 0) \equiv 0$, and $\tilde{y}_\alpha(t; h)$, $t \geq 0$, is an approximation to the zero solution, and $\tilde{Y}_\alpha(h)$ is a collection of those approximations. When we speak of a linear first approximation, it is always obtained as in (1.3). Sometimes we omit the word linear. This same equation is sometimes called the equation of first variation or simply the linearization. Let us note that when we speak of stability, it is always stability in the sense of Lyapunov and his followers; in particular, it is not numerical stability in the sense of Dahlquist.

3 Definition of $E_\alpha(t)$

Let us recall that $\alpha(h)$ is assumed to be a strictly increasing, continuous function of h for $h \geq 0$ which is zero at zero; for example, $\alpha(h)$ can be Kh^r for a positive integer r. Assume the functions $f(t, x)$ and $F(t, x)$ to be defined for $0 \leq t < \infty$ and $x \in R^d$, to be continuous in t, and locally Lipshitz in x. Besides, $F(t, 0) \equiv 0$. Assume the solution $x(t; x_0)$ of the initial value problem $\dot{x}(t) = f(t, x)$, $x(0) = x_0$, to be continuable till $t = \infty$. Only some of these assumptions are restated in theorems that follow.

The global error $e_\alpha(t; x_0; h)$ is defined as follows:

$$e_\alpha(t; x_0; h) = \sup_{\tilde{x}_\alpha \in \tilde{X}_\alpha} \|\tilde{x}_\alpha(t; x_0; h) - x(t; x_0)\|.$$

We later use this to define $E_\alpha(t)$.

Instead of saying that $\tilde{x}_\alpha(t;x_0;h)$ exists for $0 \le t \le T$, we say that $\tilde{x}_\alpha(t;x_0;h)$ can be continued till $t = T$. We say that every approximation $\tilde{x}_\alpha(t;x_0;h)$ can be continued till T if every $\tilde{x}_\alpha(t;x_0;h)$ exists for $0 \le t \le T$ with any allowed choice of discontinuities at $t = kh$. Lemma 3.1 introduces $h_0(T,r)$ and $L(T,r)$.

Lemma 3.1. *Assume, as usual, that $f(t,x)$ is continuous in t, $t \ge 0$, and locally Lipshitz in x, $x \in R^d$. Let $x(t;x_0)$, $t \ge 0$, be the unique solution of the initial value problem $\dot{x}(t) = f(t,x)$, $x(0) = x_0$. Then,*

(i) *There exists a constant $L(T,r) > 0$ such that*

$$\|f(t,x_1) - f(t,x_2)\| \le L(t,r)\|x_1 - x_2\|$$

for $0 \le t \le T$ and $\|x_i - x(t;x_0)\| \le r$, for any $T > 0$, $r > 0$, and $i = 1,2$.

(ii) *There exists a constant $h_0(T,r)$ such that for $0 < h < h_0(T,r)$ every approximation $\tilde{x}_\alpha(t;x_0;h)$ can be continued till $t = T$, and satisfies $\|\tilde{x}_\alpha(t;x_0;h) - x(t;x_0\| < r$, for $0 \le t \le T$.*

Further, for $0 < h < h_0(T,r)$, $e_\alpha(t;x_0;h) \le te^{L(T,r)t}\alpha(h)$.

Proof. Similar to proof of Theorem 3.4.6 in Stuart and Humphries [25]. Similar estimates using the Lipshitz constant are found in [9] and other places. □

For the zero solution of $\dot{y}(t) = F(t,y)$, $y(0) = 0$, where $F(t,0) \equiv 0$, $e_\alpha(t;h)$ is defined as follows:

$$e_\alpha(t;h) = \sup_{\tilde{y}_\alpha \in \tilde{Y}_\alpha} \|\tilde{y}_\alpha(t;h)\|, \tag{3.1}$$

where $\tilde{Y}_\alpha(h)$ is the family of approximations to the zero solution with time step h.

Proposition 3.2. *Assume $h < h_0(T,r)$ for $r > 0$, $T > 0$, and that $0 \le t \le T$. The $e_\alpha(t;x_0;h)$ of the solution $x(t;x_0)$ of $\dot{x}(t) = f(t,x)$, $x(0) = x_0$, and the $e_\alpha(t;h)$ of the zero solution of $\dot{y}(t) = F(t,y)$, $y(0) = 0$, where $F(t,y) = f(t,y + x(t;x_0)) - f(t,x(t;x_0))$, are the same.*

Proof. It is enough to show that members $\tilde{x}_\alpha(t;x_0;h)$ of $\tilde{X}_\alpha(x_0;h)$ and members $\tilde{y}_\alpha(t;h)$ of $\tilde{Y}_\alpha(h)$ can be matched so that $\tilde{x}_\alpha(t;x_0;h) = x(t;x_0) + \tilde{y}_\alpha(t;h)$.

It is well known (see [5]) that

$$x(t+\tau;t;x(t;x_0)+\delta) = x(t+\tau;x_0) + y(t+\tau;t,\delta)$$

for $\tau \ge 0$. Hence, if $\tilde{x}_\alpha(kh;x_0;h) = x(kh;x_0) + \tilde{y}_\alpha(kh;h)$ then $\tilde{x}_\alpha(t;x_0;h) = x(t;x_0) + \tilde{y}_\alpha(t;h)$ for $kh < t < (k+1)h$. Further, the discontinuities of $\tilde{x}_\alpha(t;x_0;h)$ and $\tilde{y}_\alpha(t;h)$ at $t = kh$, where k is a positive integer, can be exactly matched. Therefore, the approximate solutions can be matched as desired. □

The $E_\alpha(t)$ which corresponds to the zero solution of $\dot{y}(t) = F(t,y)$, $y(0) = 0$, $t \ge 0$, is defined as follows:

$$E_\alpha(t) = \limsup_{h \to 0} \frac{e_\alpha(t;h)}{\alpha(h)}. \tag{3.2}$$

By Lemma 3.1, $E_\alpha(t) \le te^{Lt}$, where $L = L(T,r)$ for some $T > t$ and $r > 0$. For a nonzero solution $x(t;x_0)$ of $\dot{x}(t) = f(t,x)$, $x(0) = x_0$, $t \ge 0$, $E_\alpha(t)$ is defined to be the same as the $E_\alpha(t)$ for the zero solution of $\dot{y}(t) = f(t,y + x(t;x_0)) - f(t,x(t;x_0))$, $y(0) = 0$, $t \ge 0$.

As in the stability theory of ODEs, we can and do confine ourselves to an analysis of the zero solution of $\dot{y}(t) = F(t, y)$ without any loss of generality. From here on, assume $L(T, r)$ to be such that

$$\|F(t, y_1) - F(t, y_2)\| \leq L(t, r)\|y_1 - y_2\|$$

for $0 \leq t \leq T$, $\|y_1\| \leq r$, $\|y_2\| \leq r$, where $T > 0$ and $r > 0$. Also, assume $h_0(T, r)$ to be such that for $0 < h < h_0(T, r)$ every approximation $\tilde{y}_\alpha(t; h)$ can be continued till $t = T$ and satisfies $\|\tilde{y}_\alpha(t; h)\| < r$.

4 Properties of $E_\alpha(t)$

This section is devoted to properties of $E_\alpha(t)$. We show that $E_\alpha(t)$ is continuous (Theorem 4.3) and independent of the choice of $\alpha(h)$ (Theorem 4.4) so that it can be written as $E(t)$. Theorem 4.5 and its corollary relate $E(t)$ to the accumulation of global error when $x(t; x_0)$ is approximated.

In the proofs in this section, we repeatedly use Theorem 10.1 of [9]. That theorem allows us to bound the divergence of two solutions of a differential equation using a Lipshitz constant.

A sequence of inequalities are often combined in a way we now describe. Let $e_0 = 0$, and $e_i \leq f e_{i-1} + r_{i-1}$ for $1 \leq i \leq n$. Then, we have,

$$e_n \leq (f^{n-1} r_0 + f^{n-2} r_1 + \cdots + r_{n-1}),$$

assuming $f \geq 0$. In fact, most of the time r_i will all be equal. In that situation, $e_n \leq r_0(1 + f + \cdots + f^{n-1})$. And when $f \geq 1$, we can get $e_n \leq f^{n-1}(r_0 + \cdots + r_{n-1})$, or $e_n \leq n f^{n-1} r_0$ if the r_i are all equal.

Lemma 4.1. *Let $h < h_0(t + s, r)$, $r > 0$. Then*

$$e_\alpha(t + s; h) \leq e^{Ls} e_\alpha(t; h) + (s + h)e^{Ls}\alpha(h),$$

where $L = L(t + s, r)$ and $s \geq 0$.

Proof. Let $\tilde{y}_\alpha(t; h)$ be an approximation to the zero solution of $\dot{y}(t) = F(t, y)$, $y(0) = 0$.

Let $t + h_l$ be the first multiple of h after t, $t + s - h_r$ the last multiple before $t + s$, and let $t + s - h_r = t + h_l + nh$. Using Theorem 10.1 of [9] and taking into account the discontinuities at multiples of h, we have

$$\begin{aligned}
\|\tilde{y}_\alpha(t + h_l; h)\| &\leq e^{Lh_l}\|\tilde{y}_\alpha(t; h)\| + h\alpha(h), \\
\|\tilde{y}_\alpha(t + h_l + kh; h)\| &\leq e^{Lh}\|\tilde{y}_\alpha(t + h_l + (k - 1)h; h)\| + h\alpha(h), \quad 1 \leq k \leq n, \\
\|\tilde{y}_\alpha(t + s; h)\| &\leq e^{Lh_r}\|\tilde{y}_\alpha(t + nh; h)\|.
\end{aligned}$$

Combining these inequalities, we get

$$\|\tilde{y}_\alpha(t + s; h)\| \leq e^{Ls}\|\tilde{y}_\alpha(t; h)\| + (s + h)e^{Ls}\alpha(h).$$

The proof is easy to complete using (3.1). □

Lemma 4.2. *With the same assumptions about h and L as in the previous lemma,*

$$e_\alpha(t + s; h) \geq e^{-Ls} e_\alpha(t; h),$$

where $s \geq 0$.

Proof. Let $\tilde{y}_\alpha(\tau; h)$, $0 \le \tau \le t$, be an approximate solution. Continue it from $\tau = t$ to $\tau = t + s$ exactly as a solution of $\dot{y}(t) = F(t, y)$. Then,

$$\|\tilde{y}_\alpha(t + s; h)\| \ge e^{-Ls}\|\tilde{y}_\alpha(t; h)\|.$$

The proof is now easy to complete using (3.1). □

Theorem 4.3. *The $E_\alpha(t)$ of the zero solution of $\dot{y}(t) = F(t, y)$, $y(0) = 0$, is continuous for* $t \ge 0$.

Proof. From Lemmas 4.1 and 4.2, we get

$$e_\alpha(t - \delta; h)e^{-L\delta} \le e_\alpha(t; h) \le e^{L\delta}e_\alpha(t - \delta; h) + e^{L\delta}(\delta + h)\alpha(h),$$

when $0 < t - \delta < t$, and when $t < t + \delta$,

$$(e_\alpha(t + \delta; h) - e^{L\delta}(\delta + h)\alpha(h))e^{-L\delta} \le e_\alpha(t; h) \le e^{L\delta}e_\alpha(t + \delta; h).$$

Divide these two inequalities by $\alpha(h)$ and use (3.2) to deduce continuity of $E_\alpha(t)$. □

Theorem 4.4. *The $E_\alpha(t)$ of the zero solution of $\dot{y}(t) = F(t, y)$, $y(0) = 0$, is the same for any $\alpha(h)$, with $\alpha(0) = 0$ and $\alpha(h)$ continuous and strictly increasing for $h \ge 0$.*

Proof. Let $\beta(h)$ be another function like $\alpha(h)$, which is continuously increasing for $h \ge 0$ and zero at zero. We will show that $E_\beta(t) = E_\alpha(t)$.

Take $L = L(t, r)$ for some $r > 0$. Take h_α and h_β small enough that the approximate solutions in the families $\tilde{Y}_\alpha(h)$ and $\tilde{Y}_\beta(h)$ are not only continuable till t but stay within a radius r of zero. For ϵ small enough, there exist unique h_α and h_β such that $\alpha(h_\alpha) = \beta(h_\beta) = \epsilon$. Assume this relationship between h_α, h_β, and ϵ. Then $h_\alpha \to 0$ and $h_\beta \to 0$ as $\epsilon \to 0$.

It is enough to show that

$$\lim_{\epsilon \to 0} \frac{|e_\alpha(t; h_\alpha) - e_\beta(t; h_\beta)|}{\epsilon} = 0.$$

In fact, it is enough to show that for every approximate solution $\tilde{y}_\alpha(t; h_\alpha)$ there is an approximate solution $\tilde{y}_\beta(t; h_\beta)$ such that

$$\|\tilde{y}_\alpha(t; h_\alpha) - \tilde{y}_\beta(t; h_\beta)\| \le \epsilon f(t)g(\epsilon), \tag{4.1}$$

where $f(t)$ is finite and $g(\epsilon) \to 0$ as $\epsilon \to 0$. It will also have to be shown that a given $\tilde{y}_\beta(t; h_\beta)$ can be approximated by some $\tilde{y}_\alpha(t; h_\alpha)$ in the same manner; but the proof of this is gotten by transposing α and β.

Given a $\tilde{y}_\alpha(t; h_\alpha)$, we construct a $\tilde{y}_\beta(t; h_\beta)$ as follows. $\tilde{y}_\beta(\tau; h_\beta)$ is determined by the differential equation $\dot{y}(t) = F(t, y)$, except when $\tau = ph_\beta$, p being a positive integer. At points $\tau = ph_\beta$, introduce a discontinuity of magnitude smaller than or equal to $h_\beta\beta(h_\beta)$ in such a way as to get as close to $\tilde{y}_\alpha(ph_\beta; h_\alpha)$ as possible.

Let $h_\beta = (m + f)h_\alpha$, $m \ge 0$ being an integer, and $0 \le f < 1$. Assume that the number of integer multiples of h_α in $(ph_\beta, (p + 1)h_\beta]$ is k_p. Let $e_p = \|\tilde{y}_\alpha(ph_\beta; h_\alpha) - \tilde{y}_\beta(ph_\beta; h_\beta)\|$.

Then $e_0 = 0$ and

$$\begin{aligned} e_{p+1} &\leq e^{h_\beta L} e_p + k_p h_\alpha \alpha(h_\alpha) e^{h_\beta L} - h_\beta \beta(h_\beta) \\ &= e^{h_\beta L} e_p + k_p h_\alpha \epsilon e^{h_\beta L} - h_\beta \epsilon. \end{aligned}$$

The first two terms can be derived from Theorem 10.1 of [9]. The third term is because of the discontinuity introduced into $\tilde{y}_\beta(\tau; h_\beta)$ at $\tau = ph_\beta$. If this bound on e_{p+1} is negative, we can take $e_{p+1} = 0$. If $n = \lfloor \frac{t}{h_\beta} \rfloor$, and there are k_n multiples of h_α in $(nh_\beta, t]$, then

$$\|\tilde{y}_\alpha(t; h_\alpha) - \tilde{y}_\beta(t; h_\beta)\| \leq e^{Lh_r} e_n + k_n h_\alpha \alpha(h_\alpha) e^{Lh_r},$$

where $h_r = t - nh_\beta$.

Let p be the highest among $\{0, 1, \dots, n\}$ such that $e_p = 0$. Combining the inequalities for e_i, $p \leq i \leq n$, we get

$$e_n \leq e^{L(n-p)h_\beta} (h_\alpha \epsilon e^{h_\beta L} (k_p + \cdots + k_{n-1}) - (n-p) h_\beta \epsilon).$$

But $k_p + \cdots + k_{n-1}$ being the number of multiples of h_α in $(ph_\beta, nh_\beta]$ is bounded above by $(n-p)h_\beta/h_\alpha + 1$. Let $t' = t - h_r$. Then,

$$\begin{aligned} e_n &\leq e^{Lt'} ((n-p) h_\beta e^{h_\beta L} \epsilon + h_\alpha \epsilon e^{h_\beta L} - (n-p) h_\beta \epsilon) \\ &\leq \max(t' e^{Lt'}, e^{Lt'}) ((e^{h_\beta L} - 1) + h_\alpha e^{h_\beta L}) \epsilon. \end{aligned}$$

Finally,

$$\|\tilde{y}_\alpha(t; h_\alpha) - \tilde{y}_\beta(t; h_\beta)\| \leq \max(e^{Lt}, te^{Lt}) ((e^{h_\beta L} - 1) + h_\alpha e^{h_\beta L} + k_n h_\alpha) \epsilon.$$

Since k_n is the number of multiples of h_α in $(t - h_r, t]$, $k_n h_\alpha < h_\beta + h_\alpha$. Thus the bound above satisfies all conditions required of (4.1), and the proof is complete. □

From here on, we drop the subscript α from $E_\alpha(t)$, $\tilde{y}_\alpha(t; h)$, $\tilde{Y}_\alpha(h)$, $\tilde{x}_\alpha(t; x_0, h)$, $\tilde{X}_\alpha(x_0; h)$, and $e_\alpha(t; h)$.

Theorem 4.5. *Let $E(t)$ be associated with the zero solution of $\dot{y}(t) = F(t, y)$, $y(0) = 0$. Given $T > 0$ and $\epsilon > 0$, there exists $h_0 > 0$ such that $h < h_0$ implies*

$$\sup_{\tilde{y} \in \tilde{Y}} \|\tilde{y}(t; h)\| = e(t; h) \leq (E(t) + \epsilon) \alpha(h)$$

for $0 \leq t \leq T$.

Further, if $\theta < E(t)$ there exists an $\epsilon > 0$ such that $e(t; h) > (\theta + \epsilon) \alpha(h)$ for arbitrarily small h.

Proof. Let $\epsilon(t; h) = e(t; h)/\alpha(h) - E(t)$.

Take $h < h_0$, where h_0 to begin with is smaller than $h_0(T, r)$ for some $r > 0$, and $L = L(T, r)$. By definition of $E(t)$ (3.2),

$$\limsup_{h \to 0} \epsilon(t, h) = 0 \tag{4.2}$$

for $0 \leq t \leq T$.

To prove the first part, given $\epsilon > 0$ we find an $h_0 > 0$ such that $h < h_0$ implies $\epsilon(t; h) < \epsilon$ for $0 \leq t \leq T$. In fact, it is enough to show that there exists an $h_t > 0$, which depends on t, and an open neighbourhood of t in $[0, T]$ such that for τ in that neighbourhood and $h < h_t$, $\epsilon(\tau; h) < \epsilon$. If the neighbourhoods of $t_1, \ldots, t_n$ give a finite cover of the compact interval $[0, T]$, we can take h_0 to be the minimum of $h_{t_1}, \ldots, h_{t_n}$.

Clearly,

$$\begin{aligned} \epsilon(t + \delta t; h) &= \epsilon(t; h) + \frac{e(t + \delta t; h) - e(t; h)}{\alpha(h)} - (E(t + \delta t) - E(t)) \\ &\leq \epsilon(t; h) + \left|\frac{e(t + \delta t; h) - e(t; h)}{\alpha(h)}\right| + |E(t + \delta t) - E(t)|. \end{aligned}$$

By (4.2), $\epsilon(t; h) < \epsilon_1$ for h small enough. By continuity of $E(t)$, $|E(t + \delta t) - E(t)| < \epsilon_2$ for $|\delta t|$ small enough.

Using Lemmas 4.1 and 4.2, and the bound $tE^{Lt}\alpha(h)$ on $e(t; h)$ from Lemma 3.1, we get

$$\frac{|e(t + \delta t; h) - e(t; h)|}{\alpha(h)} \leq \max\Big(te^{Lt}(1 - e^{-L\delta t}), te^{Lt}(e^{L\delta t} - 1) + (\delta t + h)e^{L\delta t}\Big).$$

for $\delta t > 0$, and for $\delta t < 0$,

$$\frac{|e(t + \delta t; h) - e(t; h)|}{\alpha(h)} \leq \max\Big(te^{Lt}(e^{-L\delta t} - 1), te^{Lt}(1 - e^{L\delta t}) + (\delta t + h)\Big).$$

Therefore, we can choose $|\delta t|$ and h small enough to ensure $|e(t + \delta; h) - e(t; h)|/\alpha(h) < \epsilon_3$. We need only take $\epsilon_1 + \epsilon_2 + \epsilon_3 < \epsilon$ to find the desired h_t and the neighbourhood of t.

The second half of the theorem is immediate from the definition of $E(t)$. □

Corollary 4.6. *Let $E(t)$ be associated with the solution $x(t; x_0)$ of $\dot{x}(t) = f(t, x)$, $x(0) = x_0$, where $f(t, x)$ is continuous in t and locally Lipshitz in x in an open neighbourhood of the solution $(t, x(t; x_0))$, $t \geq 0$. Given $T > 0$ and $\epsilon > 0$, there exists an $h_0 > 0$ such that*

$$\|\tilde{x}(t; x_0; h) - x(t; x_0)\| < (E(t) + \epsilon)\alpha(h)$$

for $0 \leq t \leq T$, $h < h_0$, and any approximation $\tilde{x}(t; x_0; h)$ in the family $\tilde{X}(x_0; h)$.

So far, we have defined the notion of an approximate solution (see Figure 1), and an $E(t)$ for every solution of an ODE. Theorem 4.5 and its Corollary 4.6 relate $E(t)$ to the accumulation of global error. In the upper bound $(E(t) + \epsilon)\alpha(h)$ for the global error, the details of the numerical method have been pushed into $\alpha(h)$; the conditioning function $E(t)$ is something intrinsic to the solution $x(t; x_0)$. Figure 2 shows $E(t)$ for some simple examples.

As shown in Figure 2a, the bound on global error given by $E(t)$ may be pessimistic in practice. Our model of approximations allows arbitrary discontinuities whose norms are bounded by $h\alpha(h)$, while for any given numerical method the discretization errors are fixed. Our analysis does not say how small an h is good enough for the bound on the global error to be governed by $E(t)$ as in Theorem 4.5. Will an h in a practical computation be small enough? We answer this in Henrici's words [13]: *usually if a stepsize is small enough to yield an accurate solution, it is also small enough that an asymptotic formula gives a correct indication of the size of the error.* Numerical experiments in [3] and [4] lend ample support to this statement.

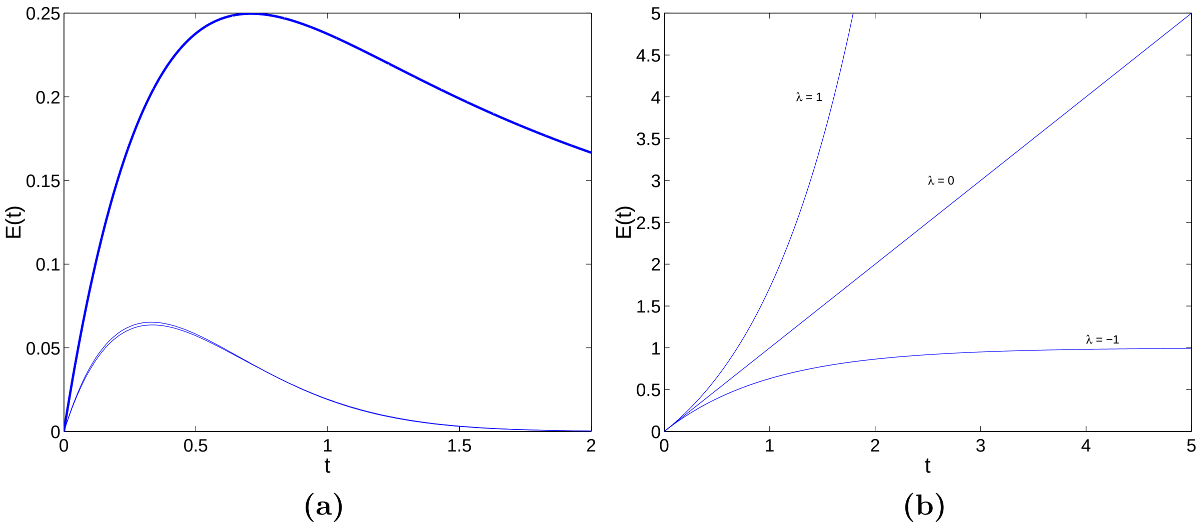

Figure 2: (a) The thick line is the $E(t)$ for $\dot{y}(t) = -(2(t+1) + (t+1)^{-1})y$, $y(0) = 1$. The two thin lines below it are global errors of forward and backward Euler divided by $8h$; here $h = .01$. (b) $E(t)$ for $\dot{y}(t) = \lambda y$. The exact $E(t)$ are obtained using Theorem 5.1.

5 $E(t)$ for Linear Systems

Our investigation of the relationship between $E(t)$ and stability properties of the exact solution begins with the linear system $\dot{y}(t) = A(t)y$, $y(0) = 0$. The relationship is not as simple as one might wish. There are both asymptotically stable examples with exponentially increasing $E(t)$ and unstable examples with linearly bounded $E(t)$. However, Theorems 5.5 and 5.6 give conditions for $E(t)$ to be bounded by a constant or to be linearly bounded. Linear systems are an important class of problems by themselves. What is more, they can be used for understanding the $E(t)$ of nonlinear systems. We will see this in Theorem 6.2 and in Section 6.2.

In $\dot{y}(t) = A(t)y$, the $d \times d$ matrix $A(t)$ is assumed to be continuous for $t \geq 0$. For such linear systems, it is easy to show that $y(t)$ is a linear function of $y(0)$ for $t \geq 0$ [22]. Thus $y(t) = Y(t)y(0)$ for $Y(t) \in R^{d,d}$, where $Y(t)$ is called the principal fundamental matrix of the linear system. The matrix $Y(t)$ itself is continuous in t and always nonsingular. Moreover, $y(t) = Y(t)Y^{-1}(s)y(s)$ for $t \geq s$. Although it does not appear to be directly related, let us mention the beautiful derivation of the Magnus series for $Y(t)$ by Iserles and Nørsett [15].

For scalar linear systems $\dot{y}(t) = a(t)y$, $a(t) \in R$, we have the following theorem.

Theorem 5.1. *The $E(t)$ for the zero solution of $\dot{y}(t) = a(t)y$, $y(0) = 0$, is given by*

$$E(t) = e^{g(t)} \int_0^t e^{-g(s)} ds,$$

where

$$g(t) = \int_0^t a(\tau) d\tau.$$

Proof. The fundamental matrix, which is scalar in this situation, is given by $Y(t) = e^{g(t)}$. Since $Y(t)$ is always positive, the optimal choice of $v(s)$ in Theorem 5.2 is $v(s) \equiv 1$. □

Let us now consider the concepts of stability put forward by Lyapunov [17]. The definitions will be stated for nonzero solutions $x(t;x_0)$.

Definition 1. The solution $x(t;x_0)$ of $\dot{x}(t)=f(t,x)$, $x(0)=x_0$, is *stable* if given any $\epsilon>0$ there exists a $\delta>0$ such that $\|x_0'-x_0\|<\delta$ implies $\|x(t;x_0')-x(t;x_0)\|<\epsilon$ for $t\geq 0$. In fact, stability implies that given $\epsilon>0$, there exists a $\delta(\tau)>0$ for every $\tau\geq 0$ such that $\|x(\tau;x_0')-x(\tau;x_0)\|<\delta(\tau)$ implies that $\|x(t;x_0')-x(t;x_0)\|<\epsilon$ for $t\geq\tau$. Let us emphasize that $\delta(\tau)$ can depend on τ.

Definition 2. The solution $x(t;x_0)$ is *asymptotically stable* if given $\epsilon>0$ there exists a $\delta(\tau)>0$ for every $\tau\geq 0$ such that $\|x_\tau'-x(\tau;x_0)\|<\delta(\tau)$ implies not only that $\|x(t;\tau,x_\tau')-x(t;x_0)\|<\epsilon$ for $t\geq\tau$ but also that $\|x(t;\tau,x_\tau')-x(t;x_0)\|\to 0$ as $t\to\infty$.

Implicit in the definitions is an assumption about the existence of solutions which begin near the solution $x(t;x_0)$. Obviously, asymptotic stability implies stability. For the scalar, linear problem $\dot{y}(t)=a(t)y$, $y(0)=y_0$, a necessary and sufficient condition for asymptotic stability is $g(t)\to-\infty$ as $t\to\infty$, where $g(t)=\int_0^t a(s)ds$. However, the following examples show that both these concepts of stability are insufficient for bounding $E(t)$.

Example 5.1. We will show using the scalar, linear problem $\dot{y}(t)=a(t)y$ that even when a solution is asymptotically stable the $E(t)$ associated with it can increase at an arbitrarily high rate. Given a rate $r(t)$, consider a continuously differentiable function $g(t)$, $t\geq 0$, $g(0)=0$, such that

1. $g(t)\leq -t$ for all $t\geq 0$,
2. $e^{g(k)}\int_0^k e^{-g(s)}ds>r(k)$ for $k=1,2,3,\ldots$

For the linear system, take $a(t)=g'(t)$. The first condition ensures asymptotic stability of the zero solution, and the second condition implies $E(k)>r(k)$ for positive integers k. Such a $g(t)$ is easy to construct. Take $g(k)=-k$ for $k=0,1,2,\ldots$. For $k-1<t<k$, $k\geq 1$, define $g(t)$ so that $g(t)\leq -t$ and

$$\int_{k-1}^{k} e^{-g(s)}ds\geq r(k)e^k.$$

This can be carried out for any continuous $r(t)$, for example $r(t)=e^t$.

Example 5.2. On the other hand, there are unstable solutions with linearly bounded $E(t)$. Consider the scalar, linear system $\dot{y}(t)=\frac{\alpha}{t+1}y$, $t\geq 0$. For this ODE, $y(t)=(1+t)^\alpha y(0)$ implying instability of the zero solution for $\alpha>0$. Yet, for $0<\alpha<1$ $E(t)$, which is $(1-\alpha)^{-1}(1+t)^\alpha((1+t)^{1-\alpha}-1)$, is linearly bounded. For $\alpha=1$, $E(t)$ is $(1+t)\log(1+t)$.

Example 5.3. In Example 5.1, $|a(t)|$ will be unbounded. Does asymptotic stability of $\dot{y}(t)=a(t)y$, $t\geq 0$, imply a linear bound for $E(t)$ if $|a(t)|$ is bounded? The answer is no; $E(t)$ can still increase exponentially in t. We sketch the construction of a $g(t)$ to show this. First take $g_1(t)=-t$ and $g_2(t)=-2t$. Take $g(t)=g_2(t)$ for $0\leq t\leq t_1$, and let $g(t)$ increase monotonically till $g(t_2)=g_1(t_2)$ for $t_2\geq t_1$, and then let $g(t)$ decrease monotonically till $g(t_3)=g_2(t_3)$ for $t_3\geq t_2$. Repeat the same construction from t_3 onwards with t_4, t_5, and t_6 in place of t_1, t_2, t_3, and so on. The construction may be arranged so that

1. if $g(\tau)=g_1(\tau)$ then $g(t)=g_2(t)$ for $f_1\tau\leq t\leq f_2\tau$ for any fixed $0<f_1<f_2<1$,

2. $|a(t)| = |g'(t)|$ is bounded.

It is easy to check that $E(\tau) \geq e^{-\tau/2}(e^{2f_2\tau} - e^{2f_1\tau})$, for τ such that $g(\tau) = g_1(\tau)$. Therefore, for $f_2 > 1/2$, $E(t)$ increases exponentially. Let us note that the linear system in this example has a negative Lyapunov exponent of -1.

Notions of stability needed for bounding $E(t)$ either by a constant or linearly are introduced after Theorem 5.2. Theorem 5.2 is comparable to Theorem 3.1 of Cano and Sanz-Serna [4].

Theorem 5.2. *The $E(t)$ of the zero solution of $\dot{y}(t) = A(t)y$, $y(0) = 0$, is given by*

$$E(t) = \sup_{v(s)} \Big\| \int_0^t Y(t)Y^{-1}(s)v(s)ds \Big\|,$$

where the supremum is over all continuous functions $v(s) : [0,t] \to R^d$ with $\|v(s)\| \leq 1$ for $0 \leq s \leq t$. As before, $Y(t)$ is the principal fundamental matrix of $\dot{y}(t) = A(t)y$ and $A(t)$ is continuous.

Proof. As usual, let the discontinuity of $\tilde{y}(t;h)$ at $t = kh$ be $h\alpha(h)v_k$. Let $n = \lfloor t/h \rfloor$ and $h_r = t - nh$. Then

$$\tilde{y}(kh;h) = Y(kh)Y^{-1}((k-1)h;h)\tilde{y}((k-1)h;h) + h\alpha(h)v_k,$$

for $1 \leq k \leq n$, and $\tilde{y}(t;h) = Y(t)Y^{-1}(nh)\tilde{y}(nh;h)$. Combine these equalities to get,

$$\tilde{y}(t;h) = \sum_{k=1}^{n} Y(t)Y^{-1}(kh)h\alpha(h)v_k, \tag{5.1}$$

This expression is also used in the proof of Theorem 6.2.

To prove,

$$\sup_{v(s)} \Big\| \int_0^t Y(t)Y^{-1}(s)v(s)ds \Big\| \leq E(t), \tag{5.2}$$

it is enough to find a $\tilde{y}(t;h)$ for every $v(s)$ such that

$$\int_0^t Y(t)Y^{-1}(s)v(s)ds = \frac{\tilde{y}(t;h)}{\alpha(h)} + \eta(h), \tag{5.3}$$

where $\eta(h) \to 0$ as $h \to 0$. Take the discontinuity of $\tilde{y}(t;h)$ at $t = kh$ to be $h\alpha(h)v(kh)$. Use (5.1) and it is apparent that $\tilde{y}(t;h)/\alpha(h)$ approximates the Riemann integral $\int_0^t Y(t)Y^{-1}(s)ds$. The standard convergence theorem for Riemann integrals gives $\eta(h) \to 0$ as $h \to 0$.

To show (5.2) in the reverse direction, it is enough to find a $v(s)$ for a given $\tilde{y}(t;h)$ so that (5.3) holds. First, consider a discontinuous $\tilde{v}(s)$ defined by $\tilde{v}(s) = v_k$ for $kh \leq s < (k+1)h$. Clearly, $\|\tilde{v}(s)\| \leq 1$. Then,

$$\int_0^t Y(t)Y^{-1}(s)\tilde{v}(s)ds = \frac{\tilde{y}(t;h)}{\alpha(h)} + \delta(h)t, \tag{5.4}$$

where $\delta(h) = \sup_{0 \leq s \leq t} \|Y(t)Y^{-1}(s+h) - Y(t)Y^{-1}(s)\|$. Since $Y(s)$ is continuous, $\delta(h) \to 0$ as $h \to 0$. Lusin's theorem [21] is a basic result for approximating measurable functions by continuous functions. It guarantees a continuous $v(s)$ such that $\|v(s)\| \leq \|\tilde{v}(s)\|$ and $\mu(v(s) \neq \tilde{v}(s)) < h$, where μ is the Lebesgue measure. If $M = \sup_{0 \leq s \leq t} \|Y(t)Y^{-1}(s)\|$, then

$$\int_0^t Y(t)Y^{-1}(s)v(s)ds = \int_0^t Y(t)Y^{-1}(s)\tilde{v}(s)ds + 2mh, \quad |m| \leq 2M. \tag{5.5}$$

Together, (5.4) and (5.5) complete the proof. □

Corollary 5.3.

$$E(t) \leq \int_0^t \|Y(t)Y^{-1}(s)\| ds.$$

Corollary 5.4. *For* $0 \leq \delta \leq t$ *,*

$$E(t) \geq \| \int_0^\delta Y(t)Y^{-1}(s) ds\|.$$

The definitions of uniform stability and uniform asymptotic stability that follow seem to have been introduced by Malkin [18]. Theorems which deduce the stability of a nonlinear system from its linear first approximation usually (always?) assume the linear first approximation to be uniformly stable or uniformly asymptotically stable [22]. The uniformity assumptions are not explicitly stated sometimes, for example in [1] and [5]. In these cases, the $A(t)$ in $\dot{y}(t) = A(t)y$ is either constant or periodic, which means that stability implies uniform stability and asymptotic stability implies uniform asymptotic stability. Uniformity assumptions are natural in the theory of Lyapunov functions as well [27]. We will find them useful for bounding $E(t)$.

Definition 3. The solution $x(t; x_0)$ of $\dot{x}(t) = f(t, x)$ is *uniformly stable* if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $\|x(\tau; x_0) - x'_\tau\| < \delta$ for $\tau \geq 0$ implies $\|x(t; x_0) - x(t; \tau, x'_\tau)\| < \epsilon$ for $t \geq \tau$.

Definition 4. The solution $x(t; x_0)$ is *uniformly asymptotically stable* if it is uniformly stable and the choice of δ in the previous definition can be made in such a way that $\|x(t; x_0) - x(t; \tau, x'_\tau)\| \to 0$ as $\tau \to \infty$ in a uniform way; i.e. given $\epsilon' > 0$ there exists $T_{\epsilon'}$ such that $\|x(t; x_0) - x(t; \tau, x'_\tau)\| < \epsilon'$ for all $t > \tau + T_{\epsilon'}$ and x'_τ satisfying $\|x(\tau; x_0) - x'_\tau\| < \delta$.

Theorem 5.5. *If the zero solution of* $\dot{y}(t) = A(t)y$*,* $y(0) = 0$*, is uniformly stable, its* $E(t)$ *is linearly bounded; i.e.,* $E(t) \leq Kt$ *for some* $K > 0$ *and* $0 \leq t < \infty$*.*

Proof. Uniform stability of the zero solution is equivalent to boundedness of $\|Y(t)Y^{-1}(s)\|$ for $t \geq s \geq 0$ [22] [27]. If $\|Y(t)Y^{-1}(s)\| \leq K$ for $t \geq s \geq 0$, Corollary 5.3 implies $E(t) \leq Kt$. □

Theorem 5.6. *If the zero solution of* $\dot{y}(t) = A(t)y$*,* $y(0) = 0$*, is uniformly asymptotically stable, its* $E(t)$ *is bounded by a constant; i.e.* $E(t) < K$ *for some* $K > 0$ *and* $0 \leq t < \infty$*.*

Proof. Uniform asymptotic stability of the zero solution is equivalent to $\|Y(t)Y^{-1}(s)\| < Me^{-\nu(t-s)}$ for $\nu > 0$, $M > 0$, and $t \geq s \geq 0$ [22] [27]. Again, we can use Corollary 5.3 to complete the proof. □

Theorem 5.5 implies that $E(t)$ for the solution of $\dot{x}(t) = Ax$, $x(0) = x_0$, is linearly bounded if all the eigenvalues of A have negative or zero real parts, and the ones with zero real part are simple. If all the eigenvalues of A have strictly negative real parts then, in fact, $E(t)$ is bounded by a constant by Theorem 5.6. The necessary stability properties of the zero solution of $\dot{x}(t) = Ax$ are verified in numerous places including [22].

6 $E(t)$ for Nonlinear Systems

This section gives two approaches to the analysis of $E(t)$ of nonlinear systems. Theorem 6.2 proves that the $E(t)$ of the solution of a nonlinear system and the $E(t)$ of the zero solution of its first approximation are the same. So one approach is to look at the linearized problem. The other approach is to directly make stability assumptions about the solution of the nonlinear system (Theorems 6.5 and 6.6). Both these approaches are illustrated in Section 7.

Lemma 6.1. *Assume that $F(t,0) \equiv 0$, that $F(t,y)$ has continuous first order partial derivatives with respect to y, and that $F(t,y)$ is continuous with respect to t, $t \geq 0$. Then,*

$$F(t,y) = A(t)y + g(t,y),$$

where $A(t) = \frac{\partial F(t,y)}{\partial y}\big|_{y=0}$ and $g(t,y) = o(\|y\|)$ as $y \to 0$ uniformly over compact intervals of t.

Proof. See [5]. □

As noted in the introduction, the following theorem is implicit in the work of Henrici [12] [13]. It might be possible to generalize it to partial differential equations.

Theorem 6.2. *As in Lemma 6.1, let $F(t,0) \equiv 0$, let $F(t,y)$ have continuous first order partial derivatives in y and be continuous in t. Then the zero solution of*

$$\dot{y}(t) = F(t,y), \quad y(0) = 0,$$

and the zero solution of

$$\dot{y}(t) = A(t)y, \quad y(0) = 0,$$

where $A(t) = \frac{\partial F(t,y)}{\partial y}\big|_{y=0}$ have the same $E(t)$.

Proof. Let $F(t,y) = A(t)y + g(t,y)$ as in Lemma 6.1. If needed, take $h < h_0(t,r)$ for some $r > 0$. To reduce clutter in the proof, denote $\tilde{y}(kh;h)$ by $\tilde{y}_k$ and $Y(kh)$ by Y_k; $Y(t)$ is the principal fundamental matrix of $\dot{y}(t) = A(t)y$. Let $n = \lfloor t/h \rfloor$ and $h_r = t - nh$.

Routine application of the variation of constants formula [9] [22] gives

$$\tilde{y}_k = Y_k Y_{k-1}^{-1} \tilde{y}_{k-1} + \int_{(k-1)h}^{kh} Y_k Y^{-1}(s) g(s, \tilde{y}(s;h))\, ds + h\alpha(h) v_k,$$

for $k = 1, 2, \ldots, n$, and

$$\tilde{y}(t;h) = Y(t) Y_n^{-1} \tilde{y}_n + \int_{nh}^{t} Y(t) Y^{-1}(s) g(s, \tilde{y}(s;h))\, ds.$$

Here, $\tilde{y}_0 = 0$ and $Y_k Y^{-1}(s) = I + O(h)$ for $(k-1)h \leq s \leq kh$. Further, $\|\tilde{y}(s;h)\| \leq te^{Lt}\alpha(h)$ for $0 \leq s \leq t$ by Lemma 3.1, and $\|g(s,y)\| = o(y)$ as $y \to 0$ uniformly over $s \in [0,t]$. Therefore, we can write $Y_k Y^{-1}(s) g(s, \tilde{y}(s;h)) = C\alpha(h)\eta(h)u(s)$ for $0 \leq s \leq t$, a constant C which can depend on t but not on h, and $\|u(s)\| \leq 1$, where $u(s) \in R^d$; here $\eta(h) \to 0$ as $h \to 0$. The expressions for $\tilde{y}_k$ and $\tilde{y}(t;h)$ become,

$$\tilde{y}_k = Y_k Y_{k-1}^{-1} \tilde{y}_{k-1} + C\alpha(h)\eta(h) \int_{(k-1)h}^{kh} u(s) ds + h\alpha(h) v_k,$$

$$\tilde{y}(t;h) = Y(t) Y_n^{-1} \tilde{y}_n + C\alpha(h)\eta(h) \int_{nh}^{t} u(s) ds.$$

Combining these equalities,

$$\frac{\tilde{y}(t;h)}{\alpha(h)} = h\sum_{k=1}^{n} Y(t)Y_k^{-1}v_k + Ch\eta(h)\sum_{k=1}^{n} Y(t)Y_k^{-1}\int_{(k-1)h}^{kh} u(s)ds$$
$$+ C\eta(h)\int_{nh}^{t} u(s)ds.$$

Both the 2nd and 3rd terms vanish as $h \to 0$ because of $\eta(h)$. Thus the nonlinear term $g(t,y)$ has no effect on $E(t)$. The proof can be formally completed using (5.1). □

Corollary 6.3. *Let $f(t,x)$ be continuous in t and have continuous first order partial derivatives with respect to x. The $E(t)$ of the solution $x(t;x_0)$ of $\dot{x}(t) = f(t,x)$, $x(0) = x_0$, and the $E(t)$ of the zero solution of $\dot{y}(t) = A(t)y$, $y(0) = 0$, where $A(t) = \frac{\partial f(t,x)}{\partial x}|_{x=x(t;x_0)}$, are the same. In fact,*

$$E(t) = \sup_{v(s)}\Big\|\int_0^t \frac{\partial x(t)}{\partial x(s)}v(s)ds\Big\|,$$

where the supremum is taken over continuous functions $v(s)$ satisfying $\|v(s)\| \leq 1$.

Proof. For a proof of the above formula for $E(t)$, note that if $Y(t)$ is the principal fundamental matrix of $\dot{y}(t) = A(t)y$, then $Y(t)Y^{-1}(s) = \frac{\partial x(t)}{\partial x(s)}$; for a proof see any basic book on ODEs. Plug this into Theorem 5.2 to get the formula. □

For comparison with Theorem 6.2, we state Theorem 33 of Chapter IX of Sansone and Conti [22]. This result is typical of Lyapunov's theory of first approximation.

Theorem 6.4. *Assume the zero solution of $\dot{y}(t) = A(t)y$, $t \geq 0$, is uniformly asymptotically stable. If $g(t,y)$ is continuous in t and y, and $g(t,y) = o(\|y\|)$ uniformly over $t \geq 0$ as $y \to 0$, then the zero solution of $\dot{y}(t) = A(t)y + g(t,y)$ is also uniformly asymptotically stable.*

The assumption about $g(t,y)$ being $o(\|y\|)$ uniformly over the semi-infinite interval $t \geq 0$ in Theorem 6.4 is quite stringent. A nonlinearity of the form $g(t,y) = t[y_1^2,\ldots,y_d^2]^T$ does not satisfy that assumption. But the theorem does not hold if the assumption is weakened to what is known about $g(t,y)$ from Lemma 6.1; for a counterexample, see Bellman [1, p. 87]. A stringent assumption about $g(t,y)$ is not required in Theorem 6.2 because how small an h we take to get convergence of $e(t;h)/\alpha(h)$ to $E(t)$ can depend upon t. In definitions of stability, in contrast, we need one small perturbation which stays "close" till $t = \infty$.

The rest of this section is about bounding $E(t)$ by making stability assumptions about the solution of a nonlinear system. For linear systems, this program was carried out in Section 4. We introduce a technique for error analysis of approximate solutions which uses Lyapunov functions. But first an example to point out some difficulties.

Example 6.1. Let us first consider the zero solution of $\dot{y}(t) = y - e^t y^3$, $y(0) = 0$, $t \geq 0$. Its $E(t)$, by Theorems 6.2 and 5.1, is $e^t - 1$. But we show that the zero solution is actually uniformly asymptotically stable. It is even exponentially asymptotically stable.

Figure 3 is the portrait of trajectories of $\dot{y}(\tau) = y - e^t y^3$. The portrait for $y \leq 0$ is a reflection about $y = 0$. Thus we can restrict ourselves to trajectories which are always in the upper half plane. Every point (t,y) with $y \geq e^{-t/2}$ is on a trajectory that is pointed downwards.

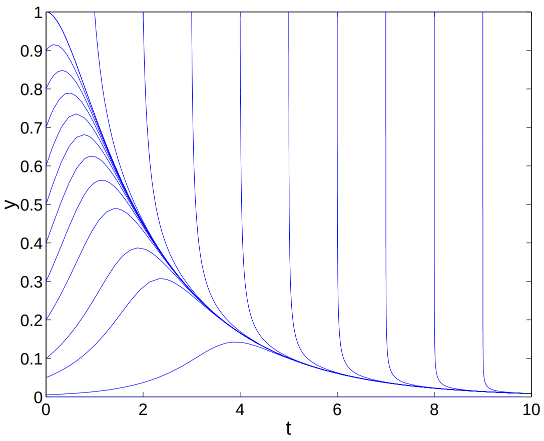

Figure 3: The portrait of trajectories of $\dot{y}(t) = y - e^t y^3$. All solutions tend towards the curve $\dot{y}(t) = e^{-t/2}$.

We now verify uniform stability of the zero solution using the last observation. Given $\epsilon > 0$, choose t_ϵ so that $e^{-t_\epsilon/2} < \epsilon$. By continuity properties, we can choose a $\delta > 0$ so that if $\tau \leq t_\epsilon$ and $y_\tau \leq \delta$, the trajectory through (τ, y_τ) stays below ϵ till t_ϵ. The maximum possible height (along y) of a trajectory beginning at (τ, y_τ), $\tau \geq t_\epsilon$ and $y_\tau < \delta$, is bounded by the larger of $e^{-\tau/2}$ and $|y_\tau|$. Since $e^{-\tau/2} < \epsilon$ and $\delta < \epsilon$, uniform stability is verified.

The verification of uniform asymptotic stability will be sketchy. We base it on the following facts:

1. The solution of $\dot{y}(t) = y - e^t y^3$, $y(0) = y_0$ tends to zero as $t \to \infty$ for $0 \leq y_0 \leq 1$,

2. Further, $y(t; y_0) \leq y(t; 1)$ for $t \geq 0$, $0 \leq y_0 \leq 1$,

3. And, $0 \leq y(t + \tau; \tau; y_0) \leq y(t; 0; y_0)$ for any $y_0 \geq 0$, $\tau \geq 0$, $t \geq 0$.

The proof of item 1 involves a bit of elementary work which we do at the end. Item 2 is trivial. For item 3, think of $y(t; 0, y_0)$ as the solution of $\dot{y}(t) = y - e^t y^3$, $y(0) = y_0$, and of $y(t + \tau; \tau, y_0) = z(t)$ as the solution of $\dot{z}(t) = z - e^{t+\tau} z^3$, $z(0) = y_0$, and use a differential inequality [11, p. 27].

Now let T_ϵ be such that $\|y(t; y_0)\| < \epsilon$ for $t \geq T_\epsilon$ and $y_0 = 1$. Then $\|y(t + \tau; \tau, y_0)\| < \epsilon$ for any $\tau \geq 0$, $t \geq T_\epsilon$ and $|y_0| \leq 1$. Thus, the zero solution is uniformly asymptotically stable.

To prove item 1, it is enough to verify that the solution with the initial condition $y(0) = 1$ satisfies $y(t) < 2e^{-t/2}$ for $t \geq 0$. This is obvious from the portrait of trajectories: every trajectory that cuts the curve $y = 2e^{-t/2}$ is in the downward direction; therefore any trajectory that starts below that curve has to stay below it for $t \geq 0$. In fact, $y(t) < 2e^{-t/2}$ for the trajectory with $y(0) = 1$ implies that the zero solution is *exponentially asymptotically stable*, if we adopt Yoshizawa's definition of exponential asymptotic stability [27].

Here we have an example which is uniformly asymptotically stable, yet has an $E(t)$ which increases exponentially with t. This is possible because the approximations $\tilde{y}(t; h)$ can introduce errors at all points kh while the definitions of stability allow just one perturbation.

The next two theorems are nonlinear analogues of Theorem 5.5 and 5.6. The proofs this time rely on the theory of Lyapunov functions. Let us note that Lyapunov functions $V(t, y)$ are always assumed to be continuous in t and locally Lipshitz in x. $V'_F(t, y)$ is the rate of increase of $V(t, y)$

along a solution of $\dot{y}(t) = F(t, y)$ which goes through (t, y). More precisely, if $y(t + \tau; t, y)$ is such a solution, then

$$V'_F(t, y) = \limsup_{\delta \to 0^+} \frac{V(t + \delta, y(t + \delta; t, y)) - V(t, y)}{\delta}.$$

If $V'_F(t, y) \leq aV(t, y)$ then $V(t+\delta, y(t+\delta; y, t)) \leq e^{a\delta}V(t, y)$, and if $V'_F(t, y) \leq 0$ then $V(t+\delta, y(t+\delta; y, t)) \leq V(t, y)$; these two facts can be inferred from differential inequalities [9] [11].

We say that $F(t, y)$ is uniformly Lipshitz in a neighbourhood of zero if $\|F(t, y_1) - F(t, y_2)\| < L\|y_1 - y_2\|$, for a constant $L > 0$ and any y_i with $\|y_1\| < r$, $\|y_2\| < r$ where $r > 0$; L is the same constant for any $t \geq 0$.

Theorem 6.5. *Let $F(t, y)$ be uniformly Lipshitz in y in a neighbourhood of 0. Assume that the zero solution of $\dot{y}(t) = F(t, y)$, $y(0) = 0$, is exponentially stable in the sense that*

$$\|y(t; t_0, y_0)\| \leq Ke^{-c(t-t_0)}\|y_0\|$$

for $\|y_0\| < r$, $t_0 \geq 0$, $t \geq t_0$, $c > 0$, and $K > 0$. Then $E(t)$ of the zero solution is bounded above by a constant.

Proof. Stability assumptions in the theorem imply existence of Lyapunov function with following properties (Yoshizawa [27, p. 97], Hale [10]):

1. $\|y\| \leq V(t, y) \leq C\|y\|$, where $C > 0$ is a constant,
2. $|V(t, y_1) - V(t, y_2)| \leq L\|y_1 - y_2\|$,
3. $V'_F(t, y) \leq -qcV(t, x)$ for some $0 < q < 1$.

The domain of $V(t, y)$ is $t \geq 0$ and $\|y\| \leq r$ for $r > 0$.

Take $h < h_0(t, r)$. Because of item 3, $V(t, \tilde{y}(t; h))$ decreases at least by a factor e^{-qch} along the approximate solution when t increases from kh to $(k + 1)h$; on that interval of t the approximate solution follows the exact solution till the discontinuity at $t = (k+1)h$. The discontinuity can cause an increase in $V(t, y)$ of at most L times its magnitude by item 2. Therefore,

$$V(kh; \tilde{y}(kh; h)) \leq e^{-qch}V((k - 1)h, \tilde{y}((k - 1)h; h)) + Lh\alpha(h),$$

for $k = 1, 2, \dots, n$ and

$$V(t, \tilde{y}(t; h)) \leq e^{-qch_r}V(nh, \tilde{y}(nh; h)).$$

Combining these inequalities, we get

$$\begin{aligned} V(t, \tilde{y}(t; h)) &\leq e^{-qch_r}L\alpha(h)\Big(\frac{1 - e^{-nqch}}{1 - e^{-qch}}\Big) \\ &\leq K\alpha(h). \end{aligned}$$

That K above can be a constant independent of h and n can be deduced from basic calculus using $h < h_0$. Now, by item 1, $\|\tilde{y}(t; h)\| \leq C\alpha(h)$ implying a constant upper bound for $E(t)$. □

The difficulty in proving Theorem 6.5 directly using the norm $\|\cdot\|$ is that when $K > 1$ the discretization error might actually be amplified by a factor greater than 1 over any given time step. Since we have to make the worst possible assumption at every time step, the final bound on $E(t)$ obtained this way will actually be exponential in t when $K > 1$. The proof of Theorem 6.5 uses a carefully constructed Lyapunov function to get around this difficulty.

Theorem 6.6. *Assume as in the previous theorem that $F(t,y)$ is uniformly Lipshitz with respect to t in a neighbourhood of $y=0$. If the zero solution of $\dot{y}(t)=F(t,y)$, $y(0)=0$, $t\geq 0$, is uniformly asymptotically stable, then $E(t)\leq Kt$ for some constant K.*

Proof. Stability assumptions in this theorem imply the existence of a Lyapunov function with the following properties: (Hale [10, Theorem 4.2, Chapter X], Yoshizawa [27])

1. $\|y\|\leq V(t,y)$,
2. $V(t,0)\equiv 0$,
3. $V_F'(t,y)\leq 0$, where $V_F'(t,y)$ is defined as in the previous proof,
4. $|V(t,y_1)-V(t,y_2)|\leq K\|y_1-y_2\|$ for some constant $K>0$.

The domain of definition of $V(t,y)$ is the same as in the previous proof.

The proof is similar to that of Theorem 6.5, but this time

$$V(kh,\tilde{y}(kh;h))\leq V((k-1)h,\tilde{y}((k-1)h;h))+Kh\alpha(h)$$

for $k=0,1,\dots,n-1$ and

$$V(t,\tilde{y}(t;h))\leq V(nh,\tilde{y}(nh;h)).$$

Combining these inequalities, we have $V(t,\tilde{y}(t;h))\leq Kt\alpha(h)$. As before, $\|\tilde{y}(t;h)\|\leq Kt\alpha(h)$, which this time implies that $E(t)\leq Kt$. □

Corollary 6.7. *Let $x(t;x_0)$ be the nonzero solution of the initial value problem $\dot{x}(t)=f(t,x)$, $x(0)=x_0$. Assume $f(t,x)$ is uniformly Lipshitz in x in a neighbourhood of $x(t;x_0)$. Then uniform asymptotic stability of $x(t;x_0)$ implies that its $E(t)$ is linearly bounded.*

Let us call attention to the necessity of making a uniform Lipshitz assumption about $F(t,x)$ or $f(t,x)$ in Theorem 6.6 or Corollary 6.7. Example 6.1, which is uniformly asymptotically stable, does not satisfy the uniform Lipshitz assumption and has an $E(t)$ which increases exponentially. We do not know if Theorem 6.6 is still true if the assumption of uniform asymptotic stability is weakened to just uniform stability. If such a theorem were true, its wider applicability might be of use. Table 1 summarizes all of Sections 5 and 6 except Theorem 6.2.

7 Three Applications to Dynamical Systems

We give three examples to illustrate the applicability of our methods for bounding the accumulation of global error.

7.1 Hyperbolic Sinks of C^1 Dynamical Systems

Let p be a fixed point of a C^1 dynamical system $\dot{x}(t)=f(x)$; i.e. let $f(p)=0$. Then p is a hyperbolic sink, if all the eigenvalues of $\frac{\partial f}{\partial x}\big|_{x=p}$ have strictly negative real parts. The following theorem can be derived using Chapter 6 of [25]. We give the theorem here because our method of proof is different.

Theorem 7.1. *Let $x(t;x_0)$ be a trajectory of the dynamical system $\dot{x}(t)=f(x)$, $f\in C^1(R^d)$, which falls into a hyperbolic sink p as $t\to\infty$. Then its $E(t)$ is bounded above by a constant.*

Linear Problems	Stability	$E(t)$ can increase exponentially even if $\|A(t)\|$ is bounded
	Asymptotic stability	$E(t)$ can increase exponentially even if $\|A(t)\|$ is bounded; Example 5.3
	Uniform stability	$E(t)$ must be linearly bounded; Theorem 5.5
	Uniform asymptotic stability	$E(t)$ must be bounded by a constant; Theorem 5.6
Nonlinear problems	Uniform stability	$E(t)$ can increase exponentially
	Uniform stability with uniform Lipshitz assumption	Not known if $E(t)$ must be linearly bounded
	Uniform asymptotic stability	$E(t)$ can increase exponentially; Example 6.1
	Uniform asymptotic stability with uniform Lipshitz assumption	$E(t)$ must be linearly bounded; Theorem 6.6
	Exponential stability as in Theorem 6.5	Not known if $E(t)$ must be linearly bounded
	Exponential stability as in Theorem 6.5 with uniform Lipshitz assumption	$E(t)$ must be bounded by a constant; Theorem 6.5

Table 1: Summary of part of Sections 5 and 6. The second column is the stability assumption about the solution; the last column says what is known about the conditioning function $E(t)$ corresponding to such a solution.

Proof. Without loss of generality, take $p = 0$. By [20, p. 150], there is a neighbourhood U_0 of 0 such that $x_0 \in U_0$ implies

$$\|x(t; x_0)\| < ce^{-at}$$

for constants $a > 0$, $c > 0$, and for $t \geq 0$.

Using continuity properties of solutions of differential equations (or openness of the basin of attraction), we can assume an open neighbourhood U of $\{x(t; x_0)\big| t \geq 0\}$ in R^d and a compact set K containing U such that K and consequently U are both contained in the basin of attraction of $p = 0$. Since all trajectories beginning in K enter U_0 in a finite amount of time, which depends only on K, we can assume

$$\|x(t; x_0)\| < Ce^{-at}$$

for constants $a > 0$, $C > 0$, for $t \geq 0$, and for any $x_0 \in K$. We can also take K and U to be invariant under the flow.

For any trajectory $x(t; x_0)$, $t \geq 0$, with $x_0 \in U$, there exists $r > 0$ such that if $\|x(t; x_0) - x_1\| < r$, $t \geq 0$ then $x_1 \in K$. For such an x_0, if $\|\delta\| < r$,

$$\|x(t; x_0) - x(t; \tau, x(\tau; x_0) + \delta)\| < 2Ce^{-at}$$

for $t \geq \tau \geq 0$. If $x_0 \notin U$ but is in the basin of attraction of p, the trajectory $x(t; x_0)$ enters U in a finite amount of time. So we can get the same kind of bound as above by adjusting C and a if necessary.

To apply Theorem 5.5 and deduce boundednes of $E(t)$, we need only verify uniformity of the Lipshitz condition on $f(x)$ for $x \in K$. This is trivial since K is compact and $f(x)$ is C^1.

□

7.2 Hyperbolic, Attracting Cycles of $C^{1+\epsilon}$ Dynamical Systems

Let $x(t)$, $t \geq 0$, be a periodic orbit of the C^1 dynamical system $\dot{x}(t) = f(x)$ in R^d. Let $T > 0$ be its period so that $x(t + T) = x(t)$. Denote the set of points on this orbit by γ.

The characteristic multipliers of the cycle γ can be defined in two ways. One is to pick a point $p \in \gamma$, take a cross-section Σ at p, define a Poincaré map for Σ, and then define the characteristic multipliers as the $(d-1)$ eigenvalues of the linearization of the Poincaré map at p. The other way is to consider the linear first approximation $\dot{y}(t) = A(t)y$ on the cycle γ. Obviously, $A(t + T) = A(t)$ for $t \geq 0$. The Floquet numbers of this linear system can also be used to define characteristic multipliers. For a lucid account of these matters, see Robinson [20].

The cycle γ is hyperbolic and attracting if all its characteristic multipliers are strictly less than 1 in magnitude.

Theorem 7.2. *Let $x(t; x_0)$, $t \geq 0$, be an orbit of a $C^{1+\epsilon}$ dynamical system $\dot{x}(t) = f(x)$ in R^d which falls into a hyperbolic, attracting cycle γ as $t \to \infty$. Then its $E(t)$ is linearly bounded from above.*

Let us first prove the following lemma. If any one solution of a linear system is uniformly stable, so is every other solution. So we might speak of the linear system itself as being uniformly stable.

Lemma 7.3. *Assume $x_0 \in \gamma$ so that $x(t; x_0)$ is a periodic orbit. Let its linear first approximation be $\dot{y}(t) = A(t)y$, $t \geq 0$. If γ is hyperbolic and attracting, $\dot{y}(t) = A(t)y$ is uniformly stable, and the $E(t)$ associated with $x(t; x_0)$ is linearly bounded.*

Proof. Uniform stability of $\dot{y}(t) = A(t)y$ is an easy consequence of the characteristic multipliers of γ being strictly less than 1. See Chapter IX of [22]. The linear bound on $E(t)$ is implied by Theorem 5.5 and Corollary 6.7. □

Lemma 7.3 is contained in a different form in the work of Cano and Sanz-Serna [4]. But let us note that Theorem 7.2 goes beyond Lemma 7.3 in a significant way. In practice, it is highly unlikely that x_0 itself is on the cycle γ. But it is often easy to find x_0 so that $x(t; x_0)$ falls into a cycle γ.

The following lemma is known as the Dini-Hukuhara-Caligo theorem. It is Corollary 1 of Chapter IX of [22]. Its proof, which we omit, is short and simple, and illustrative of an important technique in stability theory.

Lemma 7.4. *Assume the linear system $\dot{y}(t) = A(t)y$, $t \geq 0$, is uniformly stable. Assume also that $B(t)$, $t \geq 0$, is continuous with $\int_0^\infty \|B(t)\| dt < \infty$. Then the linear system $\dot{y}(t) = (A(t) + B(t))y$ is also uniformly stable.*

Proof of Theorem 7.2. By Hartman [11, p. 254], there exists a point $x_0' \in \gamma$ such that

$$\|x(t; x_0) - x(t; x_0')\| < ce^{-at}, \tag{7.1}$$

for constants $a > 0$, $c > 0$, and for $t \geq 0$. This is called convergence in phase [20].

Let $\dot{y}(t) = A(t)y$, where $A(t) = \frac{\partial f}{\partial x}\big|_{x=x(t;x_0')}$, be the first approximation along $x(t; x_0)$. By Lemma 7.3, this linear system is uniformly stable.

Let $\dot{y}(t) = (A(t) + B(t))y$, where $A(t) + B(t) = \frac{\partial f}{\partial x}\big|_{x=x(t;x_0)}$ be the first approximation along $x(t; x_0)$. The estimate (7.1) for convergence in phase implies

$$\|B(t)\| < c_1 e^{-a_1 t},$$

for constant $a_1 > 0$ and $c_1 > 0$. This is because both $x(t; x_0)$ and $x(t; x_0')$ stay within a compact region of R^d, and $f(x)$ is $C^{1+\epsilon}$ (this where we need the ϵ in $C^{1+\epsilon}$). By Lemma 7.4, $\dot{y}(t) = (A(t) + B(t))y$ is also uniformly stable.

Since $A(t) + B(t)$ gives the linearization of $x(t; x_0)$, the proof is easy to complete using Theorem 5.5 and Corollary 6.7. □

7.3 Normally Contracting Manifolds with Quasiperiodic Flows

Let us introduce the notation ϕ_t for the flow induced on R^d by $\dot{x}(t) = f(x)$. With this notation $x(t; x_0) = \phi_t x_0$. Let V be a compact C^1 manifold which is invariant under this flow. We consider the situation when the flow on V is *differentiably conjugate* to quasiperiodic flow on a torus, and V is normally hyperbolic and contracting, or briefly, *normally contracting.* We now explain the two italicized concepts in this paragraph.

A torus T^n is n copies of the circle S^1. If the angle on the ith circle is parameterized by θ_i, a quasiperiodic flow on T^n is of the form $\theta_i(t) = (\theta_i(0) + \alpha_i t) \mod 2\pi$. In fact, the flow is periodic if the α_i are all mutually commensurable. Denote this flow by ψ_t.

When we say that the flow ϕ_t is differentiably conjugate to quasiperiodic flow on a torus, we mean that there exists a C^1 homeomorphism $h : V \to T^n$ such that $h(\phi_t x) = \psi_t(hx)$ for $x \in V$ and $t \geq 0$.

To define normal contractivity [14] [20], associate a direct sum decomposition $T_x \oplus N_x$ of R^d with every x in V. In this splitting T_x is the tangent space of V at x, and N_x, the normal space, varies continuously with x. If the N_x can be chosen so that

$$\|\Pi_{N_y} \frac{\partial y}{\partial x} | N_x\| < ce^{-\mu t},$$

where $y = \phi_t x$, the matrix inside the norm is the restriction of the derivative $\frac{\partial y}{\partial x}$ to act from N_x to N_y, $c > 0$, and $\mu > 0$, then V is normally contracting. Usually, the definition of normal contractivity comes with an other assumption which says contraction in the normal direction dominates any contraction on the manifold V. But since we have assumed that the flow on V is differentiably conjugate to quasiperiodic flow on a torus, this other assumption can be dropped.

Obviously, the tangent spaces T_x are invariant under the derivative map $\frac{\partial \phi_t x}{\partial x}$. It is actually possible to choose N_x so that they too are invariant under the derivative map [14] [20]. We take this to be the case. So the derivative map $\frac{\partial \phi_t x}{\partial x}$ maps T_x to $T_{\phi_t x}$ and N_x to $N_{\phi_t x}$.

Theorem 7.5. *Let $x(t; x_0)$ be a trajectory of the $C^{1+\epsilon}$ flow $\dot{x}(t) = f(x)$ which falls into a normally contracting and invariant manifold V. Assume that the flow on V is differentiably conjugate to quasiperiodic flow on a torus. Then the $E(t)$ of $x(t; x_0)$ is linearly bounded.*

The above theorem generalizes Theorem 7.2. Its proof is exactly analogous. We begin with a lemma about trajectories that begin on V.

Lemma 7.6. *Let $x_0 \in V$ so that $x(t; x_0)$ stays on V for $t \geq 0$. Make the same assumptions about V as in Theorem 7.5. Then the linearization $\dot{y}(t) = A(t)y$ along $x(t; x_0)$ is uniformly stable, and the $E(t)$ of $x(t; x_0)$ is linearly bounded.*

Proof. The principal fundamental matrix of the linearization in the lemma is given by $Y(t) = \frac{\partial \phi_t x_0}{\partial x_0}$, which is the derivative map. Therefore, it is enough if we show that $\|\frac{\partial \phi_t x}{\partial x}\|$ is bounded by a constant for any $x \in V$ and $t \geq 0$.

We already know that the maps induced by the derivative map between tangent spaces and between normal spaces are bounded in norm because of differentiable conjugacy to flow on a torus and normal contractivity, respectively. Since the tangent spaces and normal spaces are both invariant under the derivative map, it is enough if we show that the angle between T_x and N_x (in the sense of the CS decomposition) is bounded away from 0. That this angle is bounded away from 0 is implied by the compactness of V. □

The proof of Theorem 7.5 can be completed exactly like the proof of Theorem 7.2 using the following result about convergence in phase.

Theorem 7.7. *As in Theorem 7.2, let $x(t; x_0)$ be a trajectory of the C^1 flow $\dot{x}(t) = f(x)$ which falls into a normally contracting and invariant manifold V, and let the flow on V be differentiably conjugate to quasiperiodic flow on a torus. Then there exists $x_0' \in V$ such that*

$$\|x(t; x_0) - x(t; x_0')\| < ce^{-at},$$

for positive constants c and a, and $t \geq 0$.

Proof. This theorem can be deduced from Theorem 4.1 and the remark following its proof in Hirsch, Pugh and Shub [14]. See in particular part (a) of that theorem about stable manifolds and part (g) about conjugacy to linearized flows. □

8 Three Remarks

This last section is a collection of parenthetical remarks about matters which are related to $E(t)$ and which we have not investigated in detail.

(i) *Multistep methods and variable time stepping.* The model for discretization errors in Section 2 is adapted to single step methods with constant step sizes. For linear multistep methods with constant step sizes, we believe the accumulation of global error can be worse but not better (after excluding some trivial cases) than indicated by $E(t)$.

For variable time stepping, it is sometimes but not always true that the ratio of the largest to the smallest time step is bounded by a constant [23] [24]. In this case at least, the effect of variable time stepping is to improve global errors by a constant factor over the indication given by $E(t)$.

(ii) *One sided Lipshitz conditions.* Let us briefly summarize the approach to global error analysis using one sided Lipshitz conditions for the linear system $\dot{y}(t) = A(t)y$. Since $\|y(t)\|^2 = y^T(t)y(t)$, we have

$$\frac{d\|y(t)\|^2}{dt} = y^T(t)(A^T(t) + A(t))y(t).$$

If $\lambda(t)$ is the maximum eigenvalue of $A^T(t) + A(t)$, then

$$\frac{d\|y(t)\|^2}{dt} \leq \lambda(t)\|y(t)\|^2.$$

Thus upper bounds for $\|y(t)\|$, and hence for $\|Y(t)\|$ where $Y(t)$ is the principal fundamental matrix of the linear system, can be written down in terms of $\lambda(t)$. These can be plugged into Theorem 5.2 to get bounds on $E(t)$ and hence the accumulation of global error. For a detailed account, see [9].

In our view, one sided Lipshitz conditions are basically a way to get a handle on stability by looking at the evolution of the norm of $y(t)$. This is a far less general approach to stability than the two methods of Lyapunov we have used (this deficiency of norms is something Lyapunov must have realized). It is also of far lesser applicability; we do not see a way to derive any of the results in Section 7 by using one sided Lipshitz conditions.

(iii) *Numerical symplectic integrators and Hamiltonian problems.*

Special methods which preserve properties like symplecticity or energy conservation or volume conservation in phase space show milder increase of global errors with t than general methods in numerically integrating Hamiltonian problems. Calvo and Sanz-Serna [3] showed using careful numerical experiments that the accumulation of global error with t was linear for a symplectic method but quadratic for a Runge-Kutta method when the methods were applied to Kepler's problem. A full explanation came after [2] and [7] in the paper [4] by Cano and Sanz-Serna. There are numerical experiments in Quispell and Dyt [19] which are unexplained.

The conditioning function $E(t)$ is not suitable for studying the accumulation of global error in these special methods. The class of local discretization errors allowed by the model in Section 2 is too general. The discretization errors made by these special methods are restricted in some way; for example, the discretization errors of an energy preserving method cannot take the approximation out of the constant energy manifold. But is it possible to define conditioning functions which are adapted to these special methods? We think it might be possible. The concepts of stability needed to bound such conditioning fuctions will also, no doubt, be different.

9 Acknowledgements

It is a pleasure to thank Arieh Iserles and Mike Shub for encouragement and enlightening discussions. I have benefited from talking to several visitors to M.S.R.I., Berkeley, in Fall 98 about this work. I thank Timo Eirola and Andrew Stuart for suggestions which directly influenced the presentation of this paper.

References

[1] R. Bellman, *Stability Theory of Differential Equations*, McGraw-Hill, New York, 1953.

[2] M.P. Calvo and E. Hairer, Accurate long-term integration of dynamical systems, *Appl. Numer. Math. 18* (1995), 95-105.

[3] M.P. Calvo and J.M. Sanz-Serna, The development of variable-step symplectic integrators, with application to the two-body problem, *SIAM J. Sci. Comput. 14* (1993), 936-952.

[4] B. Cano and J.M. Sanz-Serna, Error growth in the numerical integration of periodic orbits, with application to Hamiltonian and reversible systems, *SIAM J. Numer. Anal. 34* (1997), 1391-1417.

[5] E.A. Coddington and N. Levinson, Theory of Ordinary Differential Equations, McGraw-Hill, New York, 1955.

[6] G. Dahlquist, Stability and error bounds in the numerical integration of ordinary differential equations, *Trans. of the Royal Inst. of Techn.,* Stockholm, Sweden, Number 130.

[7] D.J. Estep and A.M. Stuart, The rate of error growth in Hamiltonian-conserving integrators, *Z. Agnew. Math. Phys. 46* (1995), 407-418.

[8] W.B. Gragg, *Repeated extrapolation to the limit in the numerical solution of ordinary differential equations*, Thesis, Univ. of California, 1964; see also *SIAM J. Numer. Anal. 2* (1965), 384-403.

[9] E. Hairer, S.P. Nørsett and G. Wanner, *Solving Ordinary Differential Equations I*, Springer-Verlag, New York, 1980.

[10] J. Hale, *Ordinary Differential Equations*, John Wiley, New York, 1969.

[11] P. Hartman, *Ordinary Differential Equations*, John Wiley, New York, 1973.

[12] P. Henrici, *Discrete Variable Methods in Ordinary Differential Equations*, John Wiley, New York, 1962.

[13] P. Henrici, *Error Propogation for Difference Methods*, John Wiley, New York, 1963.

[14] M.W. Hirsh, C.C. Pugh and M. Shub, *Invariant Manifolds*, Lecture notes in Mathematics 583, Springer-Verlag, New York, 1977.

[15] A. Iserles and S.P. Nørsett, On the solution of linear differential equations in Lie groups, *Phil. Trans. of Royal Soc. A*, to appear.

[16] A. Iserles and G. Söderlind, Global bounds on numerical error for ordinary differential equations, *J. Complexity 9* (1993), 97-112.

[17] A. M. Lyapunov, *Problème générale de la stabilité du mouvement*, Comm. Soc. Math. Kharkov 2 (1892), 3 (1893); Ann. Fac. Sci. Toulouse 9 (1907), 204-474; Ann. of Math. Studies 17, Princeton University Press, Princeton, 1949.

[18] J. G. Malkin, *Theory of Stability of Motion*, Atomic Energy Commision Tech. Rep. 3352, 1956.

[19] G.R.W. Quispel and C.P. Dyt, Volume-preserving integrators have linear error growth, *Physics Letters A 242* (1998), 25-30.

[20] C. Robinson, *Dynamical Systems. Stability, Symbolic Dynamics, and Chaos*, CRC press, Boca Roton, 1995.

[21] W. Rudin, *Real and Complex Analysis*, 3rd Edition, McGraw-Hill, New York, 1987.

[22] G. Sansone and R. Conti, *Non-linear Differential Equations*, Macmillan, New York, 1964.

[23] D. Stoffer and K. Nipp, Invariant curves for variable step size integrators, *BIT 31* (1991), 169-180.

[24] A.M. Stuart, Probabilistic and deterministic convergence proofs for software for initial value problems, *Numerical Algorithms 14* (1997), 227-260.

[25] A.M. Stuart and A.R. Humphries, *Dynamical Systems and Numerical Analysis*, Cambridge University Press, Cambridge, UK, 1996.

[26] L.N. Trefethen and D. Bau III, *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[27] T. Yoshizawa, *Stability Theory by Liapunov's Second Method*, Mathematical Society of Japan, 1966.

Mathematical Sciences Research Institute (MSRI)
1000 Centennial Drive
Berkeley, CA 94720.
divakar@cs.cornell.edu